EXPERIENCE WITH BAYE'S THEOREM FOR COMPUTER DIAGNOSIS OF CONGENITAL HEART DISEASE*

Homer R. Warner,† Alan F. Toronto, L. George Veasy
Department of Biophysics and Bioengineering, University of Utah and Latter-day Saints Hospital, Salt Lake City, Utah

The problem of computer diagnosis is an example of the general problem of pattern recognition. It is the purpose of this paper to describe a mathematical model of medical diagnosis fundamentally based on Baye's Theorem but incorporating certain expansions and modifications which were added as the result of experience in applying the model to the diagnosis of congenital heart disease.[1-3] A computer program has been written which permits translation of statistical data concerning the incidence of symptoms in diseases into a prediction of the probability of particular patients having particular diseases. Refinements in the mathematical model of diagnosis have been introduced as the result of experience in the application of this program to a patient population. On the other hand, experience with the program has yielded insight into the intricacies of diagnosis in a particular field of medicine and has resulted in improved accuracy of diagnosis, both by the computer program and the participating physicians who feed patient data to the computer and who review the computer's diagnoses.

The expanded form of Baye's Theorem, which forms the basis of this mathematical model and has been used for the diagnosis of congenital heart disease in this laboratory over the past two and one-half years, is shown in:

$$A_{j,S} = \prod_{i=1}^{M} (P_{i,j})^{a_i} (1 - P_{i,j})^{b_i} (EX_{i,j})^{c_i} \tag{1}$$

and:

$$P(j/S) = \frac{A_{j,S}}{\sum_{j=1}^{N} A_{j,S}} \tag{2}$$

$P_{i,j}$ is a matrix of dimensions M by N, where i identifies a symptom and j a disease. $P_{i,j}$ is the probability of a patient with the jth disease having the ith symptom. In this study, the symptoms include age group, physical findings, phonocardiographic abnormalities, electrocardiographic abnormalities, and certain abnormalities seen in a chest X-ray, $P_{i,}$ is the *a priori* incidence

*This work supported by Grant FR-00012 from the Division of Research Facilities and Resources of the National Institutes of Health, and a grant from The American Heart Association.

†Recipient of Research Career Award (HE-18-344-02) from the National Heart Institute, National Institutes of Health.

of the *jth* disease in the population of patients having one of the diseases considered by the program. Each patient may have only one disease as shown by:

$$\sum_{j=1}^{N} P_{1,j} = 1.0 \tag{3}$$

The term EX is explained later in regard to the handling of mutually exclusive symptoms. S is any particular array of M digits, each of which may be either 1, 0, or -1 and are labeled s_i. If, in the patient under study, the *ith* symptom is present, absent or uncertain in the opinion of the observer, s_i is declared 1, 0, or -1, respectively, by the doctor who fills out a check-off list from his examination of the patient. The exponents a_i and b_i in Equation 1 are determined for a given patient by the value assigned to s_i as shown in:

$$a_i = (s_i + |s_i|)/2 \tag{4}$$

and:

$$b_i = 1 - |s_i| \tag{5}$$

$P_{j,S}$ is the calculated probability that the patient characterized by the array of symptoms S has the *jth* disease. This calculation is carried out by the program for each disease and results in a differential diagnosis with a probability assigned to each disease. In the program for diagnosis of congenital heart disease, M (the number of symptoms) is 53 and N (the number of diseases) is 35.

Certain assumptions are made in this mathematical model of medical diagnosis. First, it is assumed that the symptoms are independent of one another within a given disease. This is expressed in:

$$P_{i,j}(s_k = 1) = P_{i,j}(s_k = 0) \tag{6}$$

which states that the probability of a patient with this *jth* disease presenting with the *ith* symptom is independent of whether the patient has the *kth* symptom $(s_k = 1)$ or has not $(s_k = 0)$ the *kth* symptom where k may be any integer from 1 to M except i. For a given disease matrix this assumption can be tested by chi square analysis if sufficient data are available on the coincidence of symptoms in each disease. Through the application of this model, however, it has been found that diagnosis can be improved if the set of symptoms is not limited by this constraint, but certain mutually exclusive symptoms are used as well. The special means for handling such symptoms will be described.

The second basic assumption is that the diseases are mutually exclusive, as expressed in Equation 3.

If a patient is found to have more than one disease, this new combination of diseases must then be considered as a new disease entity by expanding

the disease-symptom matrix to include it. This may not be possible if the disease is so rare that sufficient statistics cannot be accumulated. Of course, the clinician who must try to diagnose this disease without sufficient experience is faced with the same limitation.

It is not possible to calculate the probability of a disease combination from data on the incidence of symptoms in each of the separate diseases. To illustrate this point, consider the two diseases, pulmonary stenosis and atrial septal defect. Since neither of these diseases alone will produce cyanosis, it would be impossible to predict just from incidence of cyanosis in the independent lesions that a patient with both lesions might present with cyanosis. Thus, pulmonary stenosis plus atrial septal defect must be considered a new disease entity.

Finally, it is assumed that the true incidence (for which the matrix $P_{i,j}$ is an estimator) of the symptoms in each of the diseases is stationary, i.e., it is not changing with time. If, in a particular field of application, the true incidence does change with time, then the matrix $P_{i,j}$ must be updated at periods which are short relative to the period between the fluctuations in the symptom pattern of the diseases in question. For instance, the symptoms of pneumonia are dependent on the etiological agent, and the characteristics of this agent may change as the pattern of antibiotic therapy changes. In the present study this was not a problem, but frequent changes in $P_{i,j}$ were made to improve the accuracy of this estimator since the original matrix used was based on an insufficient number of cases to accurately represent the true incidence in the case of certain diseases. The effect of this "learning" process will be evident from the results shown below.

In TABLE 1 is shown the check-off list to be completed by the clinician after examination of each patient. The murmurs are evaluated from a phonocardiographic tracing* recorded from right second intercostal space (ICS) along the sternum, second left ICS, fourth left ICS and at the apex. The presence or absence of the X-ray symptoms are determined from a single posterior-anterior chest film. The clinician is asked to mark those symptoms which are present with a 1 and mark a -1 by those to be omitted, either because he is uncertain as to its presence or because the information is not available to him, i.e., no X-ray on the patient. s_i, of course, is always 1.

The brackets enclose mutually exclusive symptoms and not more than one symptom in such a set may be present. The rules for handling sets of mutually exclusive symptoms may be illustrated by referring to symptoms 5, 6, 7, and 8 which represent various forms of cyanosis. If a patient presents with any one of these, he cannot, by definition, have one of the others. Thus, it would be a mistake to use the complement of the probability of mild cyanosis as an additional piece of information if the patient presents with severe cyanosis since this necessarily excludes the mild form. In terms

*Elema Schonander Model No. 21C, which writes directly with a jet of ink.

TABLE 1
SYMPTOM CHECK LIST

1. *A priori* incidence	30. Post systolic
Age	31. Post continuous
2. Less than 1 year	32. *Murmur louder than gr 3/6 (10 mm.)
3. 1 year to 20 years	33. Accentuated P_2
4. 20 or more years	34. Diminished P_2
From physical examination	35. Fixed split P_2
5. Cyanosis, mild	**ECG findings**
6. Cyanosis, severe (with clubbing)	36. Atrial fibrillation or broad notched P wave
7. Cyanosis, intermittent	37. Axis, right (more 110°)
8. Cyanosis, differential	38. Axis, left (less than 0°)
9. Squatting	39. R wave greater than 1.2 mv in lead V_1
10. Femoral pulse less than Brachial	40. rR' or qR in lead V_1
From phonocardiogram	41. R wave greater than 2.5 mv in lead V_6
11. Apex systolic	42. T wave inversion in lead V_6
12. Apex systolic, holo	**X-ray findings**
113. Apex systolic, mid	43. Rib notching
14. Apex diastolic	44. Peripheral vessels increased
15. Apex diastolic, early	45. Peripheral vessels decreased
16. Apex diastolic, late	46. Hilar vessels increased
17. L 4th systolic	47. Hilar vessels decreased
18. L 4th systolic, holo	48. Main pulmonary artery large
19. L 4th systolic, mid	49. Main pulmonary artery not seen
20. L 4th continuous	50. Aorta large
21. L 4th diastolic	51. Aorta small
22. L 4th diastolic, holo	52. Cardiomegaly
23. L 4th diastolic, early	53. Snowman
24. L 2nd systolic	
25. L 2nd systolic, holo	
26. L 2nd systolic, mid	
27. L 2nd continuous	
28. R 2nd systolic	
29. R 2nd diastolic	

*10 mm. deflection of phonocardiogram at 1/5 sensitivity.

of the mathematical model, if symptom 5 is present, $s_5 = 1$, $s_6 = -1$, $s_7 = -1$, $s_8 = -1$, but if symptoms 5 through 8 are absent (no cyanosis), then s_5 through s_8 are -1, C_5 becomes 1 and the appropriate value for EX is obtained from a table. In this case:

$$EX_{5,j} = 1 - P_{5,j} - P_{6,j} - P_{7,j} - P_{8,j} \quad (7)$$

This represents the probability of a patient with the *jth* disease having no cyanosis. The exponent C_i is zero except when all of a set of mutually exclusive symptoms are absent.

Still a different situation is illustrated by the case of symptoms 11, 12, and 13. A systolic murmur may be further classified according to its time course of intensity into midsystolic or holosystolic in most cases. If this subclassification in a given case cannot be made with confidence, it is better not attempted even though this subclassification of the patient's murmur potentially has more separating power, i.e., contains more information. As here defined, the probability of the two subtypes of systolic murmurs and if no systolic murmur is present, the term s_{11} is set to 0 by the observer and s_{12} and s_{13} are set to -1 by the program. This approach provides the physician with a means for resorting to a reliable but less specific piece of information if he is unsure of which of the more detailed classifications to use.[4] This has been found to improve diagnostic accuracy.

For the past 18 months we have been applying these particular equations to the diagnosis of congenital heart disease using a symptom-disease matrix consisting of 53 symptoms and 35 disease entities. The list of diseases (TABLE 2) does not include all possible combinations of congenital defects. Instead, it is limited to defects and combinations of defects on which some statistics were available from which to make an initial symptom-disease matrix $(P_{i,j})$. As better statistics are accumulated, the number of diseases will be expanded but it has been held constant at 35 over this last 18-month period to permit the evaluation presented here. As shown in TABLE 1, the 53 symptoms include 25 heart murmurs and other criteria determined from a phonocardiographic tracing, 7 EKG findings, 11 X-ray findings, six other findings from physical examination, and three age group categories. No data from heart catheterization or dye injection studies were used.

Eighty-three patients were seen by each of two examining physicians over this period of time. The cases were arbitrarily divided into those seen before September 1962, and those seen after this time in order to permit an assessment of the diagnostic performance of both the physicians and the computer with time. Two statistics were used for assessing the performance of both the computer and the physicians.[3] The first of these was the fraction of cases in which the computer or physician made the correct diagnosis, as determined from follow-up studies. This means that a probability of at least .01 was assigned to that disease. The second measure of diagnostic performance was the average probability rating given to the correct diagnosis

TABLE 2
DISEASE LIST

1. Normal	18. Patent ductus arteriosus
2. Atrial septal defect	19. Pulmonary arterio-venous Fistula
3. Atrial septal defect with pulmonary stenosis	20. Congenital mitral disease
4. Atrial septal defect with pulmonary hypertension*	21. Primary myocardial disease
5. Atrio-ventricular communis	22. Anomalous origin of coronary artery
6. Partial anomalous pulmonary venous connection	23. Congenital aortic disease
7. Total anomalous pulmonary venous connection	24. Ventricular septal defect with pulmonary flow $\leq$ 1.4 systemic flow
8. Tricuspid atresia (without transposition)	25. Coarctation of aorta
9. Ebstein's anomaly	26. Truncus arteriosus
10. Ventricular septal defect with valvular pulmonary stenosis	27. Transposition
11. Ventricular septal defect with infundibular pulmonary stenosis	28. Hypertrophic subaortic stenosis
12. Pulmonary stenosis, valvular, gradient $\geq$ 40 mm. Hg.	29. Absent aortic arch
13. Pulmonary stenosis, infundibular, gradient $\geq$ 40 mm. Hg.	30. Ventricular septal defect with pulmonary flow $>$ 1.4 systemic flow
14. Pulmonary atresia	31. Ventricular septal defect with pulmonary hypertension*
15. Peripheral pulmonary stenosis	32. Patent ductus arteriosus with pulmonary hypertension*
16. Pulmonary hypertension*	33. Tricuspid atresia with transplantation
17. Aortic pulmonary window	34. Pulmonary stenosis gradient $<$ 40 mm. Hg.
	35. Ruptured sinus Valsalva

*Mean pulmonary artery pressure $\geq$ mean systemic artery pressure.

by the computer or by the physician. Each physician after completing his examination of a patient, including the chest X-ray, phonocardiogram, and electrocardiogram, and after transferring this information to the symptom check-off lists, records his own differential diagnosis of the patient. He lists a probability rating by each disease in his differential in space provided on the back of this form. This differential diagnosis is later used as a basis for comparison of the physician's diagnosis with the differential diagnosis provided by the computer program from the same symptoms. In FIGURE 1 is shown a comparison of the computer and physician using the cases seen prior to September 1962. The probability rating given to the correct diagnosis by the computer using the symptoms supplied by the physician and the most recent data matrix (here referred to as the new data matrix) is plotted

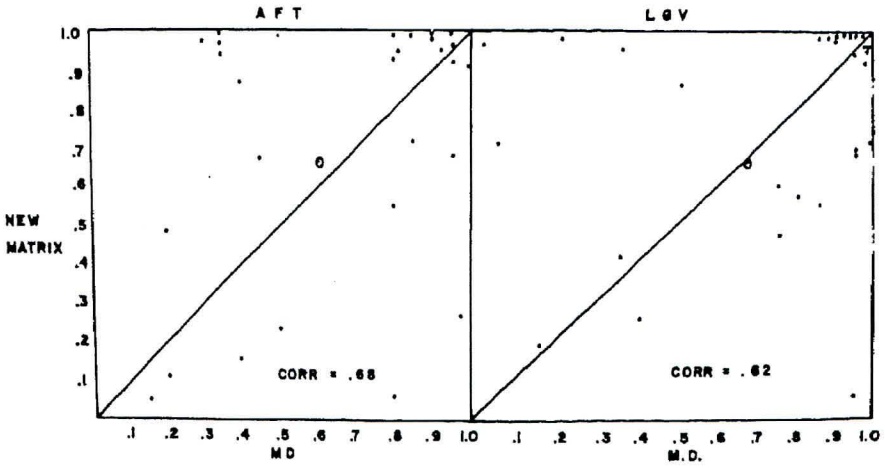

FIGURE 1. Cases seen before September 1962. Comparison of probability rating given to the correct diagnosis by physician *(abscissa)* and computer program *(ordinate)* using observations provided by the physician *(AFT* or *LGV)* against whom it is compared.

against the probability assigned to the correct diagnosis by the doctor. AFT is a clinical physiologist and LGV is an experienced pediatric cardiologist. Notice a small cluster of points near the zero coordinates representing those cases which were missed by both the physican and the computer. The clustering of cases in the upper right corner are cases in which both the computer and physician gave a high probability to the right diagnosis. Cases in the region above the line of identity were diagnosed better by the computer than by the physician and below this line, better by the physician than by the computer. On the average, the computer did better than AFT (large circle) and as well as LGV.

Using the case material seen in this laboratory since September 1962, a similar comparison was made (FIGURE 2). The results on these most recent

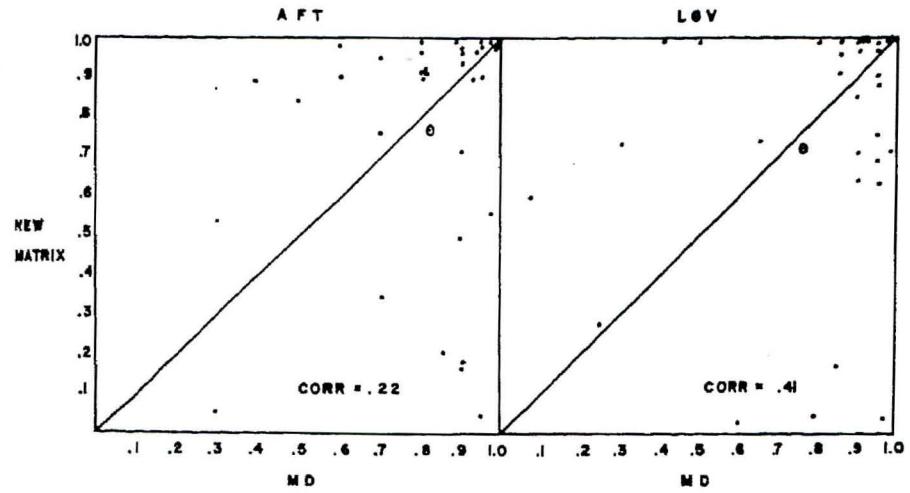

FIGURE 2. Cases seen after September 1962. Comparison as in FIGURE 1 using most recent cases.

cases show that there is a significant improvement by both criteria in the performance of both the computer and the physician. Both physicians improved more than the computer program. Notice that the observer AFT does not misdiagnose any of these 40 cases nor does the computer using the information supplied by AFT, since there are no points at the zero coordinate. It is possible that the improved performance by the computer program is due entirely to its being given more reliable observational data by the physician.

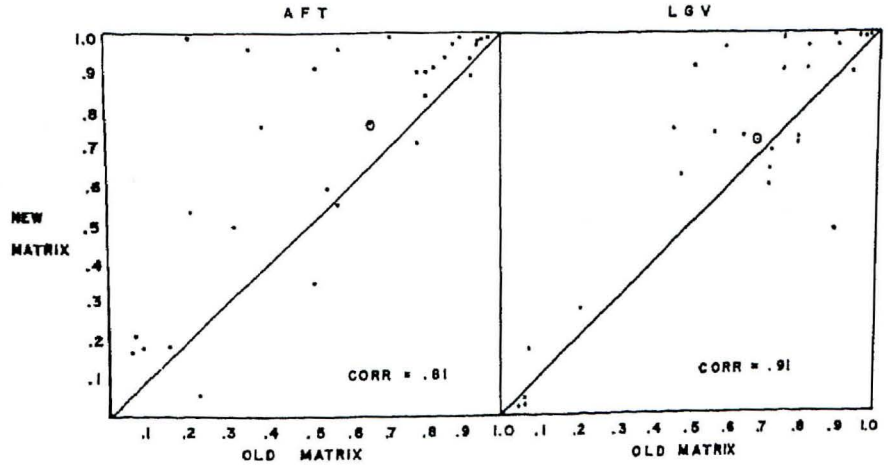

FIGURE 3. Cases seen after September 1962. Comparison of probability rating given to the correct diagnosis by the computer program using the old and the new matrix.

On the other hand, some of this improvement may be due to the fact that the symptom-disease matrix has been progressively improved as additional statistics are accumulated for each disease. To evaluate this, a comparison was made of the computer performance using a data matrix containing the accumulated experience up to February 14, 1963 (referred to as "old matrix") with the computer's performance using the new data matrix which includes experience up to May 15, 1963.

In this comparison (FIGURE 3), of course, the input data for each patient is the same whether the program uses the old or the new matrix. The only difference between the two conditions is the statistical information in the data matrix. Notice that there are many more points lying on the upper left-hand side of the line of identity than there are below this line indicating that the improved computer performance, with time, is due in part at least,

TABLE 3
DIAGNOSTIC PERFORMANCE INDEX = $\bar{P}^* \times F^\dagger$

Number of cases	Before Sept. 1962 43	After Sept. 1962 40	Average 83 (total)
A F T	0.53	0.82	.66
Old matrix	0.41	0.66	.52
New matrix	0.58	0.77	.66
L G V	0.61	0.74	.67
Old matrix	0.49	0.65	.55
New matrix	0.56	0.71	.63

*$\bar{P}$ = Mean probability assigned to correct diagnosis.
†F = Fraction of cases in which correct diagnosis was given probability > 0.01.

to the improvements in the statistical matrix and gives hope that further experience will lead to further improvement. This should be particularly true in the case of those more uncommon diseases on which adequate statistics are not yet available.

An index of diagnostic performance may be defined as the product of the two criteria already described, namely: the average probability assigned to the correct diagnosis ($\bar{P}$) and the fraction of cases (F) in which the correct disease was given a rating of at least 0.01 in the differential diagnosis. Such an index should be a better measure of diagnostic performance than $\bar{P}$ or F alone since it is sensitive to both the average rating and the complete failures. Examination of TABLE 3 shows the marked improvement in performance of observer AFT and the somewhat less dramatic improvement in observer LGV. The computer improved using both the old data matrix and the new data matrix but its performance was significantly better with the new matrix, both on the old cases and on the new case material. The

extent to which the improved performance of the physicians is the result of experience in preparing data for and receiving feedback from the computer over this period of time is difficult to evaluate. It is interesting in this regard that observer AFT, who improved the most, had the most direct contact with the computer results over the period of this study. Of course, this contact in any given case occurred after the observers had committed their differential diagnoses to writing and had assigned their probability estimates to each of the diseases in their differential diagnosis.

Summary

An equation based on Baye's Theorem has been described which has been programmed for a general-purpose digital computer for medical diagnosis. A consideration of the assumptions inherent in this approach was presented along with the methods for handling exceptions, such as mutually exclusive symptoms. Experience with this approach in the field of congenital heart disease indicates that these diseases can be diagnosed with an accuracy equal to that of an experienced specialist in this field. Furthermore, the accuracy of the computer diagnosis is still improving with refinements in the data matrix. It is suggested that perhaps a part of the improvement in the diagnostic accuracy of the physicians in this study may be attributed to their experience in preparing data for the computer and receiving feedback from the computer in the form of a differential diagnosis.

References

1. LEDLEY, R. S. & L. B. LUSTED. 1959. The use of electronic computers to aid in medical diagnosis. Proc. I.R.E. **47:** 1970.
2. WARNER, H. R., A. F. TORONTO, L. G. VEASY & R. STEPHENSON. 1961. A mathematical approach to medical diagnosis. J. Am. Med. Assoc. **177:** 177.
3. TORONTO, A. F., L. G. VEASY & H. R. WARNER. 1963. Evaluation of computer program for diagnosis of congenital heart disease. Progr. Cardiovascular Diseases **5:** 362.
4. WARNER, H. R. 1961. Medical diagnosis using a digital computer. Proc. Third IBM Med. Symp. **3:** 293.